



RAIL HEALTH SCORE

XXXXXX Data Science



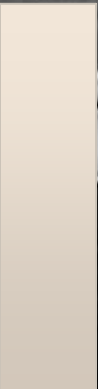
[REDACTED] ([REDACTED] subsidiary of
[REDACTED])

[REDACTED]



DATA SCIENCE

Challenges



Model Description

Model Type	>	<p><u>Version 1</u> uses direct probability modeling to predict the likelihood of a locomotive suffering from a fault in the future.</p> <p><u>Version 2</u> uses tree logic to predict the likelihood of a locomotive <i>and its various subsystems</i> suffering from a fault in the future.</p>
Inputs	>	
Outputs	>	
Value	>	

Model Description

Model Type	>	<p><u>Version 1</u> uses direct probability modeling to predict the likelihood of a locomotive suffering from a fault in the future.</p> <p><u>Version 2</u> uses tree logic to predict the likelihood of a locomotive <i>and its various subsystems</i> suffering from a fault in the future.</p>
Inputs	>	<p><u>Version 1</u> takes streaming data from locomotive sensors, such as temperatures, pressures, and electrical current.</p> <p><u>Version 2</u> expands the input to include historical locomotive repair data, age of the locomotive, and locomotive model type.</p>
Outputs	>	
Value	>	

Model Description

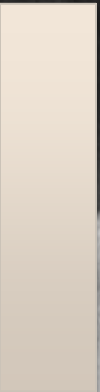
Model Type	>	<p><u>Version 1</u> uses direct probability modeling to predict the likelihood of a locomotive suffering from a fault in the future.</p> <p><u>Version 2</u> uses tree logic to predict the likelihood of a locomotive and its various subsystems suffering from a fault in the future.</p>
Inputs	>	<p><u>Version 1</u> takes streaming data from locomotive sensors, such as temperatures, pressures, and electrical current.</p> <p><u>Version 2</u> expands the input to include historical locomotive repair data, age of the locomotive, and locomotive model type.</p>
Outputs	>	<p><u>Version 1</u> outputs the overall health score of a locomotive at a given time, ranging from 0 to 100.</p> <p><u>Version 2</u> outputs the overall health score as well as the health scores of 5 locomotive subsystems at a given time, ranging from 0 to 100.</p>
Value	>	

Model Description

Model Type	>	<p><u>Version 1</u> uses direct probability modeling to predict the likelihood of a locomotive suffering from a fault in the future.</p> <p><u>Version 2</u> uses tree logic to predict the likelihood of a locomotive and its various subsystems suffering from a fault in the future.</p>
Inputs	>	<p><u>Version 1</u> takes streaming data from locomotive sensors, such as temperatures, pressures, and electrical current.</p> <p><u>Version 2</u> expands the input to include historical locomotive repair data, age of the locomotive, and locomotive model type.</p>
Outputs	>	<p><u>Version 1</u> outputs the overall health score of a locomotive at a given time, ranging from 0 to 100.</p> <p><u>Version 2</u> outputs the overall health score as well as the health scores of 5 locomotive subsystems at a given time, ranging from 0 to 100.</p>
Value	>	<p><u>Version 1</u> enables analysts to prioritize assets of the most common locomotive model type (SD70ACe) that are unhealthy, reducing resource demands significantly.</p> <p><u>Version 2</u> generates health scores across all locomotive model types and provides insight into which subsystem of an unhealthy locomotive needs attention.</p>

DATA SCIENCE

How XXXXXX Helped



Overview

How do we define a “fault”?

- A “fault” is defined as a symptom or warning of an imminent failure
- The health score addresses “control system faults,” or faults severely detrimental to a locomotive’s operations as selected by our client, YYY (YYY)

Overview

How do we define a “fault”?

- “Fault” defined as symptom/warning of an imminent failure
- Health score addresses “control system faults:
 - Or faults severely detrimental to a locomotive’s operations as selected by YYY

How do we define a “health score”?

- Health score model calculates probability of a “negative health event” (defined above) occurring in the next 2 weeks, ranging from 0 to 1.
- “Health score” computed by subtracting probability calculated above from 1, and multiplying by 100.
- “Health score” is defined as percentage probability that asset will have no major issues in the next two weeks.

Overview

How do we define a “fault”?

- “Fault” defined as symptom/warning of an imminent failure
- Health score addresses “control system faults”
 - Or faults severely detrimental to locomotive’s operations as selected by YYY

How do we define a “health score”?

- Health score model calculates probability of “negative health event” (defined above) occurring in the next two weeks, ranging from 0 to 1

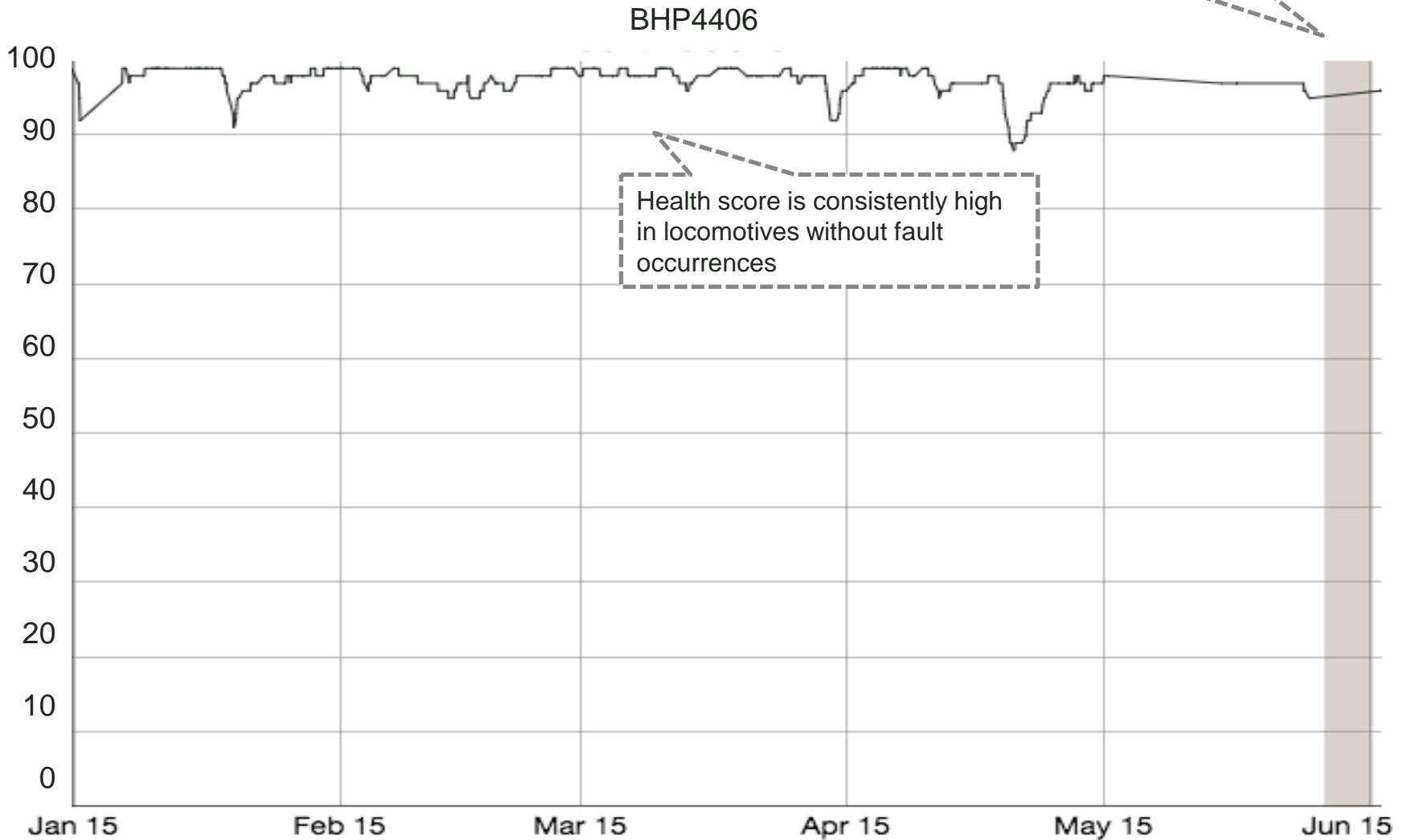
“Health score” computed by subtracting probability from 1, multiplying by 100

- In other words, “health score” is defined as the percentage probability that an asset will have no major issues in the next two weeks

How are the subsystems defined?

- Defined to us by YYY

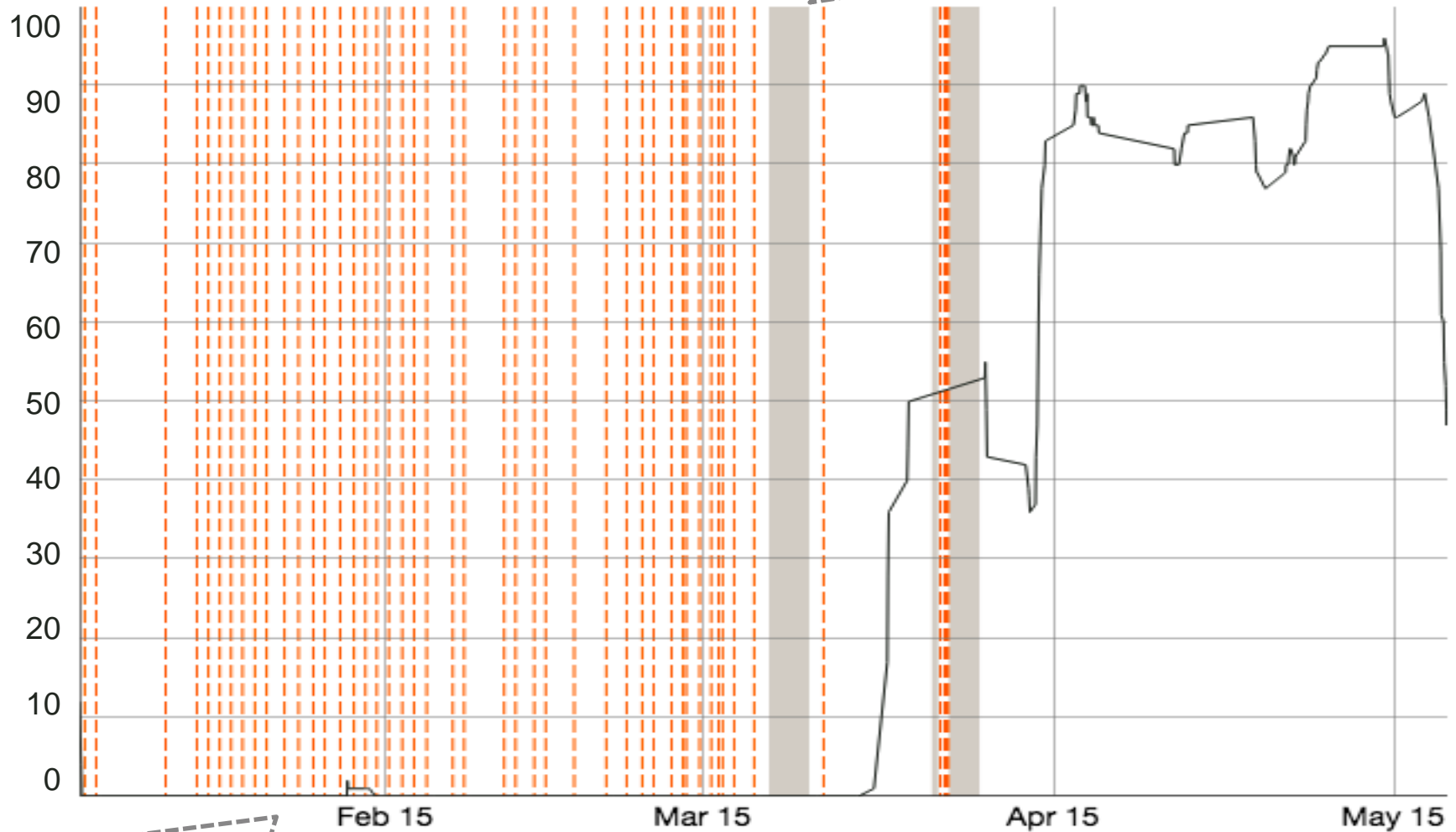
Example of a healthy locomotive



Example of an unhealthy locomotive

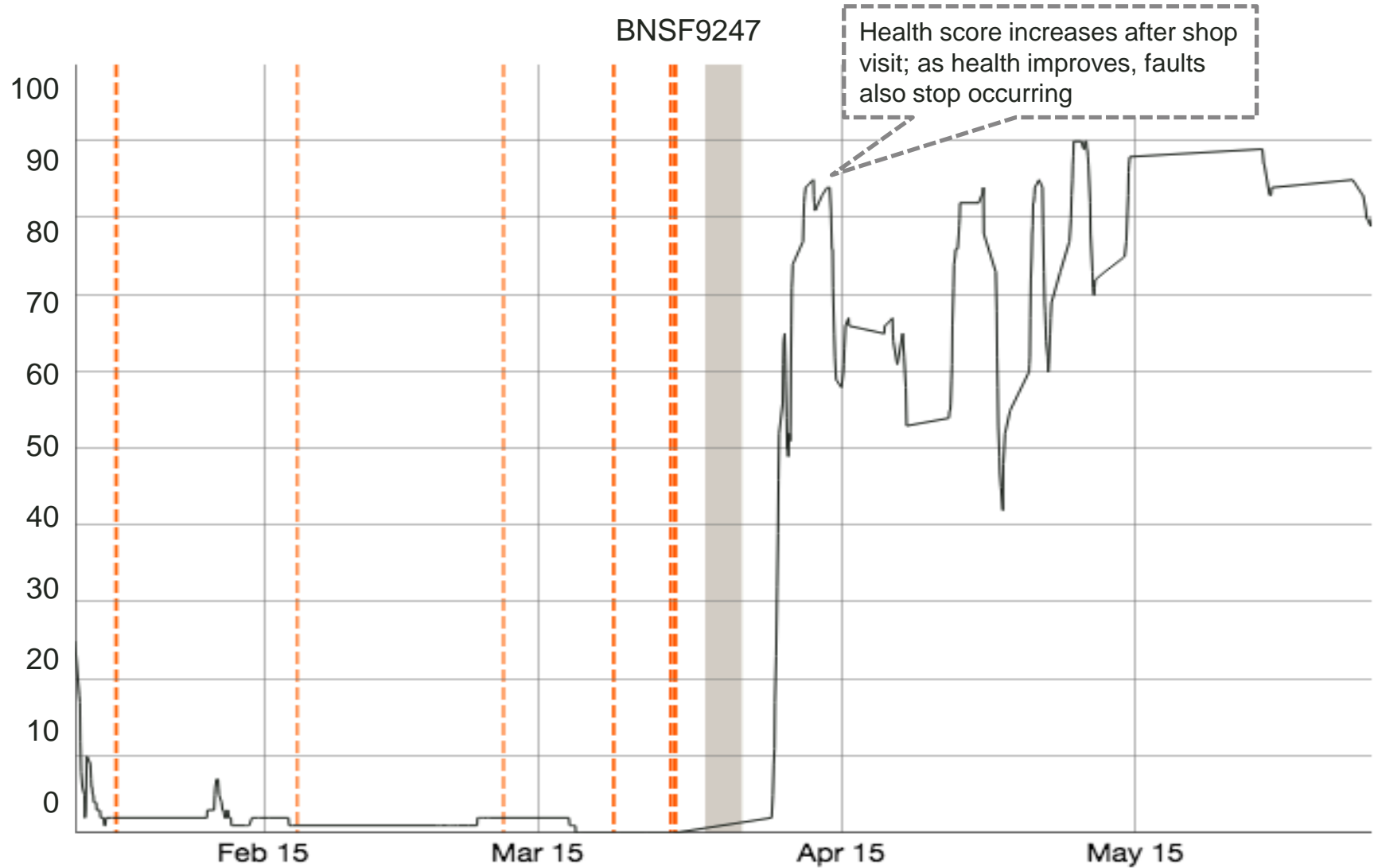
NS1075

Dotted vertical lines correspond to a fault at that time.



Health score is near 0 when faults occur frequently

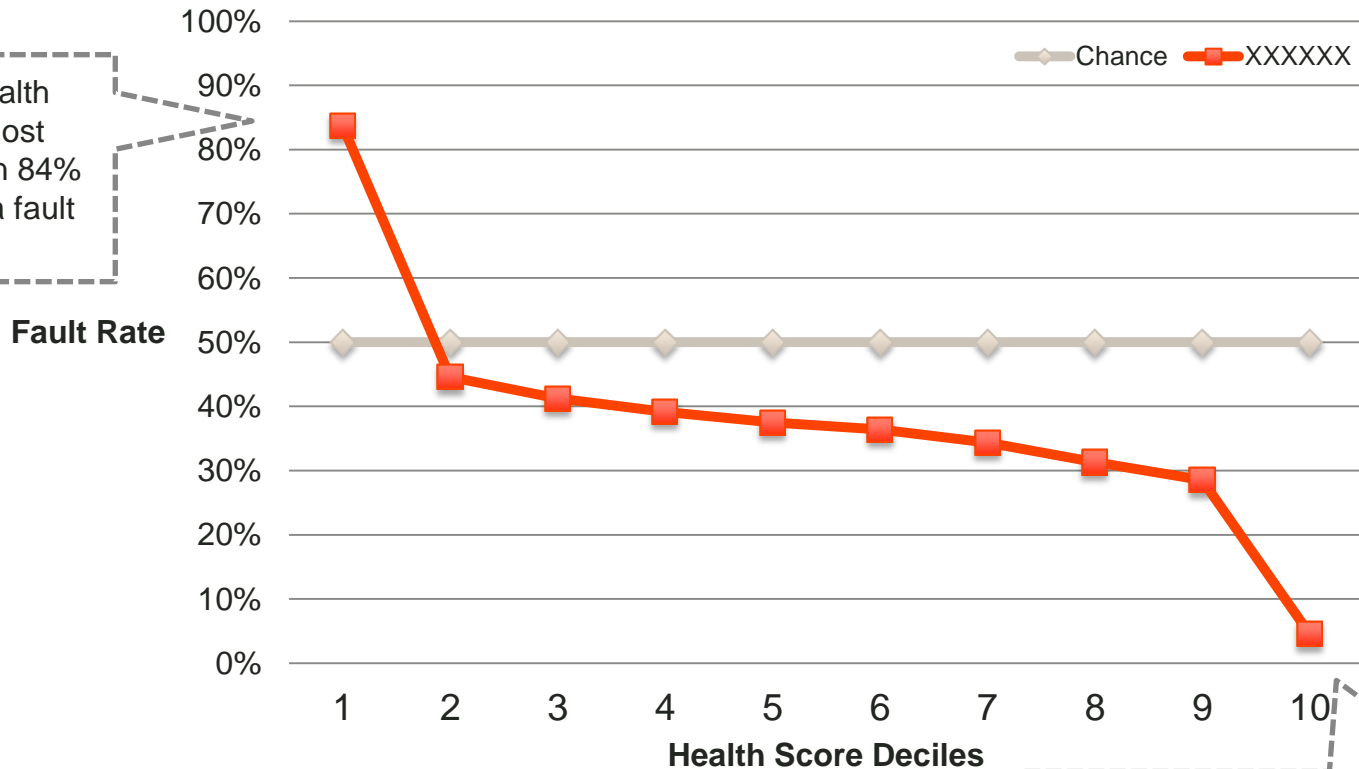
Example of an asset with increasing health



Performance

How can we evaluate the performance of our health score?

Health Score Rank Order



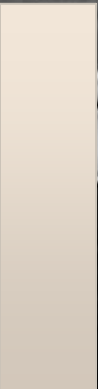
Locomotives with health scores in the lowermost decile (0-10) have an 84% chance of incurring a fault in the next 2 weeks

The XXXXXX health score model improves predictive power significantly over leaving predictions to chance.

For the highest decile (91-100), that number is just 5%!

DATA SCIENCE

YYY Rail Health Score Results



Roadmap for Rail Health Score

- Our goal:
 - Generate a score that accurately represents health of a locomotive
 - Analysts take corrective measures before locomotive breaks down
- What we've achieved:
 - Create Locomotive Health Score – predict the probability of one/more key faults occurring within a 2-week window
 - Create Subsystem Health Score – identify subsystem of unhealthy locomotive requiring most attention
- Immediate next steps:
 - Incorporate weather data – determine how meteorological factors (precipitation, terrain, temperature) affect health of locomotive
 - Link health scores to failures – predict probability of related failure given particular subsystem health score